THE EFFECT OF FATIGUE ON
JUDGMENTS OF INTERPRODUCT SIMILARITY

Working Paper #579

Michael D. Johnson
Daniel R. Horne
The University of Michigan
Donald R. Lehmann
Columbia University

FOR DISCUSSION PURPOSES ONLY

ABSTRACT

Similarity scaling often requires subjects to produce such a large number of judgments that subject fatigue may become a problem. Yet it is unclear whether or how respondent fatigue affects similarity judgments. The present study supports the notion that subjects adopt simplified representations and produce less complex judgments as they progress through a similarity rating task.

INTRODUCTION

Marketing researchers apply a variety of similarity scaling techniques, including multidimensional scaling (Shepard 1962; Kruskal 1964), hierarchical clustering (Johnson 1967), additive clustering (Shepard and Arabie 1979), and additive tree scaling (Sattath and Tversky 1977), to help understand consumer perceptions of product or service alternatives (cf. Arabie, Carroll, DeSarbo and Wind 1981; Cooper 1983; Green and Carmone 1970; Johnson and Fornell 1987; Srivastava, Leone and Shocker 1981). The underlying assumption in these applications is that perceptions, or subjects' cognitive representations of products, are an important input to consumer judgment and choice.

Similarity scaling requires a reasonable number of products to construct meaningful representations (Klahr 1969). At the same time, similarity scaling is often limited by the number of products that can be included (Hauser and Koppelman 1979). Asking subjects to make too many judgments may affect the quality of the judgments (Sudman and Bradburn 1982). Consider traditional applications in which respondents are asked to provide similarity ratings of product pairs. Because the number of ratings required for the analysis increases roughly as the square of the number of items (i.e., $n(n-1)/2$), applications involving a large number of products or services become prohibitive.

Although marketing researchers understand that there is some limit to the amount of "quality" information that can be collected from respondents, the effects of fatigue are unclear. One possible reaction is for subjects to adopt simple product representations and provide less complex judgments. Alternatively, fatigue may result in carelessness and an increase in error variance. After briefly describing how

repetition and fatigue may affect similarity judgments, we present a
study that examines these competing predictions.

TASK REPETITION, ADAPTATION, AND FATIGUE

The traditional method for collecting similarity scaling data is
to have respondents rate the overall similarity of each possible pair of
products on a proximity scale. A number of studies have examined the
reliability of these direct similarity judgments and obtained mixed
results (cf. Day, Deutscher, and Ryans 1976; Moore and Lehmann 1982;
Summers and MacKay 1976; Weksel and Ware 1967). Nevertheless,
similarity scaling, in particular multidimensional scaling (MDS),
appears fairly robust to changes in a number of factors, including the
metric employed (Green 1975), the order of presentation of the stimuli
(Jain and Pinson 1976), and the embedding of stimuli in a stimulus
domain (Malholtra 1987).

At the same time, marketing researchers recognize the often
burdensome and boring nature of similarity judgment tasks and the
fatigue that may result (Malholtra 1987). Yet it is unclear just how
respondents adapt to fatigue. Dong (1983) found that while missing and
inconsistent responses tend to increase over time, fatigue does not
appear to influence aggregate similarities or MDS solutions. A problem
with Dong's study, however, is that it may be difficult to detect
changes in individual judgments from aggregate data.

At least two factors may affect an individual's judgments through
the course of a similarity rating task: adaptation and fatigue.
Initially, respondents adapt to thinking about the items involved.
Variance in the use of the scale should decrease, leading to higher
test-retest correlations. Further, the basis of the judgments

2

themselves, whether common or distinctive features (Tversky 1977) or product categorizations (Rosch 1975), should become well established. With early repetition, respondents should "settle in" and provide more consistent judgments. We expect that this adaptation either occurs relatively quickly or is minimized, if not eliminated, by appropriate task procedures (e.g. a warm-up task or prior acclamation to the stimulus set).

As respondents continue to progress through the task, there is a danger that fatigue may affect the judgments. One possible reaction is that respondents adopt simple representations and incorporate less information into their judgments in order to "finish the task." Only the most salient differences or similarities among the alternatives may affect the subjects' responses. In judging soft-drinks, for example, consumer may begin by distinguishing alternatives on flavor, sweetness, and calories and, as the task drags on, end up distinguishing the soft-drinks almost exclusively on flavor (e.g. cola versus noncola). Alternatively, subjects may simply make judgments more haphazardly or carelessly resulting in greater judgment variance.

A decrease in judgment complexity may be reflected at a surface level by changes in the mean and variance of the subjects' ratings. The stimuli may begin to appear either very similar or very different. The subjects' responses may polarize and/or exhibit greater or lower variance over time. Whether judgment variance increases or decreases should depend on the nature of the stimuli. The more inherently similar the stimuli, such as brands from a well-defined product category, the more subjects may come to rely on salient commonalities. The more dissimilar the stimuli, such as categories themselves, the more subjects

3

may come to rely on salient differences. As a result, brand-based judgments may exhibit less variance over time as subjects focus on salient commonalities, while product category judgments may exhibit more variance as subjects focus on salient differences.

It is possible, however, that any decrease in judgment complexity is only evident at a deeper level, such as the ability of a similarity scaling technique to capture perceptions. Scaling solutions based on judgments collected early in the judgment task should exhibit more complexity and higher stress for a given number of dimensions than solutions based on judgments collected late in the judgment task. Whether brands or categories are involved, fewer aspects should be considered causing a decrease in the stress of a scaling solution.

A simple alternative prediction is that, over time, fatigue results in carelessness and an increase in the error inherent in similarity judgments. While average similarities may not change, judgment variance should increase along with the stress of MDS representations, independent of the brand or category nature of the stimuli. The following study tests these competing predictions.

STUDY DESIGN

The data used in the study was obtained as part of a larger study in which pair-wise similarity judgments were collected for five stimulus sets: soft-drinks, candy bars, beverages, snack foods, and lunch products. These stimuli were relevant for the student subjects used to provide data. They also represent two different levels of abstraction or generality: brand-level competitors from the same basic-level category (soft-drinks and candy bars) and superordinate category

alternatives from similar though different categories (beverages, snack foods, lunch products). These stimulus sets are presented in Table 1.

Each stimulus set contained 12 product alternatives requiring subjects to make 66 paired comparison ratings. Each subject rated all 66 product pairs for one of the five stimulus sets. A total of 24, 24, 24, 24, and 27 subjects (total n=123) rated the soft-drinks, candy bars, beverages, snack foods, and lunch products respectively. Half of the subjects in each group rated the 66 pairs in one random order and the other half rated the same 66 pairs in the reverse order. All pairs were rated on an 11-point similarity rating scale ranging from 0 (Very Dissimilar) to 10 (Very Similar). Subjects were run through the task in small groups (approximately 20 per group) and were led through the instructions by an experimenter. The instructions included a list of the twelve products the subject would be rating along with a sample similarity judgment scale. Overtly exposing the subjects to the range of products in the task and the similarity scale should help minimize any adaptation to the task. Any major changes in the similarity judgments should, as a result, be due to fatigue.

ANALYSIS

The prediction that perceptions simplify over time was tested by comparing "first half" ratings, based on the subjects' first 33 pairs rated, with "second half" ratings, based on the subjects' second 33 pairs rated. Means and standard deviations were calculated for the first half and second half of each subject's similarity ratings. Each half was then scaled using nonmetric multidimensional scaling in two dimensions. The stress of the two-dimensional solution (Kruskal 1964) provides a measure of fit. (Two dimensions seemed to best represent the

5

responses in most cases.) Again, if subjects adopt simpler representations over time, scaling solutions based on first half data should be more complex and exhibit higher stress than solutions based on second half data. One might also expect changes in the means and standard deviations of the responses, with the direction of a change dependent on the brand or category nature of the stimuli. If, alternatively, subjects simply become careless and error prone, judgment variance and stress should increase irrespective of the stimuli involved.

Analysis of variance models tested for changes in the means and standard deviations of the similarity judgments as well as the stress of the two-dimensional solutions. The independent variables included the first v. second half order of the judgments (two levels), the brand v. category level of the stimuli (two levels), a half by level interaction, and random effects variables for the individual stimulus sets (nested within each stimulus level) and the order conditions (nested within each stimulus set).

To provide a finer grade analysis, the judgments were also broken into sets of 22 first, second, and third-third judgments and examined. We shall concentrate, however, on the half-level data and analysis for three reasons. First, the robustness of MDS solutions has been shown to vary the more incomplete the data (Malholtra et al., 1988). Second, the results of the three-level data analysis were very consistent with the two-level analysis. Third, the fit of the two-level analytical models dominated the fit of the three-level models.

RESULTS

We first examine the means and standard deviations of the judgments presented in Table 2. There were no significant main effects on mean similarity for first v. second half or stimulus level. There was a marginally significant half by level interaction ($F=3.19$, $p<.10$). The brand stimuli became slightly less similar and the category stimuli became slightly more similar over time. However, there were no differences in average first half v. second half judgments within any of the five individual stimulus sets. Thus one obvious indicator of change in respondents behavior, the average judgment, suggests no drastic changes occur.

However, the standard deviations of the judgments varied significantly both by half ($F=6.22$, $p<.05$) and level ($F=25.48$, $p<.001$). Judgment variance was greater for the category- than the brand-level stimuli, which is natural given their greater inherent heterogeneity. More importantly, there was a significant half by level interaction ($F=19.76$, $p<.001$). As predicted, judgment variance decreased from the first to the second half judgments for the brand-level stimuli while it increased for the category-level stimuli.

This result supports the notion that subjects adopt simpler representations as they progress through a similarity judgment task. The decreasing variance in the brand-based judgments suggests that the subjects came -to rely on the brands' salient commonalities. The increasing variance in the category-based judgments suggests that the subjects came to rely on the categories' salient differences. The observed interaction is inconsistent with the notion that subjects simply became careless or error prone in their responses.

The analysis of the stress measures also supports the notion that representations simplify through the course of a judgment task. There was a very significant decrease in stress from first to second half judgments ($F=138.854$, $p<.001$). The average stress was .102 and .050 respectively for the first half and second half input. There was no difference in stress for brands v. categories. There was, however, a significant half by level interaction effect on stress ($F=6.025$, $p<.05$). The decrease in stress over time was slightly greater for the brands than for the categories. At the same time, each of the five individual stimulus sets exhibited a significant decrease in stress from the first half to second half judgments.

CONCLUSIONS

Overall the results demonstrate the systematic effect of fatigue on ratings of product similarity. The observed reduction in stress from first half to second half similarities supports a reduction in the complexity of respondents' product representations and judgments over time. The observed interaction between first v. second half judgments and stimulus level in affecting judgment variance also supports this conclusion. Judgment variance increased for the category-level stimuli while it decreased for the brand-level stimuli. This suggests that subjects relied increasing on the more salient properties of the stimuli.

There are at least two important implications of these results for marketing research. First, basic changes in how subjects think about products and rate similarity may not be obvious to the casual observer. While all five of the product categories studied showed a significant decrease in stress from first half to second half ratings, none

exhibited significant differences in average similarity. Second, the results suggest that judgments simplify rather than degenerate over time. Subjects simply came to rely on the salient properties of the judgment stimuli. At least for the task studied here, in which subjects provided 66 paired comparison ratings, fatigue did not result in greater error. Rather, the longer the judgment task, the greater is the relative weight of the stimuli's most salient similarities or differences in forming similarity judgments.

TABLE 1

STIMULUS SETS

BRAND-LEVEL STIMULI

Soft-Drinks	Candy Bars
Sprite	Three Musketeers
Seven-Up	Mars Bar
Diet Sprite	Milky Way
Diet Seven-Up	Snickers
Orange Crush	M&M Plain
Diet Orange Crush	M&M Peanut
Coke Classic	Hershey's Plain
New Coke	Hershey's Almond
Pepsi	Nestle's Crunch
Cherry Coke	Reece's Peanut Butter Cups
Diet Coke	Twix Carmel
Diet Pepsi	Kit Kat

CATEGORY-LEVEL STIMULI

Beverages	Snacks	Lunch Products
Ice Cream Soda	Popcorn	Carrot
Milk Shake	Nacho Chips	Apple
Chocolate Milk	Crackers	Fruit Juice
Milk	Potato Chips	Yogurt
Fruit Juice	Cheese	Milk
Lemonade	Grapes	Ice Cream
Soft-Drink	Apple	Cookie
Diet Soft-Drink	Yogurt	Candy Bar
Club Soda	Ice Cream	Soft-Drink
Iced Tea	Cookie	Pizza
Bottled Water	Candy Bar	Chicken Sandwich
Iced Coffee	Brownie	Hamburger

TABLE 2

Differences Between First and Second Half
Paired Comparison Similarities

BRAND-LEVEL STIMULI

Stimuli	Mean	Standard Deviation	Stress
Soft-Drinks:			
First Half	3.659	2.840	.100
Second Half	3.201	2.094	.043
Candy Bars:			
First Half	4.056	2.135	.112
Second Half	3.918	1.892	.038

CATEGORY-LEVEL STIMULI

Stimuli	Mean	Standard Deviation	Stress
Beverages:			
First Half	4.086	2.380	.109
Second Half	4.409	2.616	.057
Snacks:			
First Half	3.915	2.664	.091
Second Half	4.048	2.688	.063
Lunch Products:			
First Half	3.566	2.531	.098
Second Half	3.804	2.686	.049

REFERENCES

Arabie, Phipps, J. Douglas Carroll, Wayne DeSarbo, and Jerry Wind (1981), "Overlapping Clustering: A New Method for Product Positioning," Journal of Marketing Research, 18 (August), 310-317.

Cooper, Lee G. (1983), "A Review of Multidimensional Scaling in Marketing Research," Applied Psychological Measurement, 7 (Fall), 427-450.

Day, George S., Terry Deutscher, and Adrian B. Ryans (1976), "Data Quality, Level of Aggregation, and Nonmetric Multidimensional Scaling Solution(s)," Journal of Marketing Research, 13 (February), 92-97.

Dong, Hei-Ki (1983), "Method of Complete Triads: An Investigation of Unreliability in Multidimensional Perception of Nations," Multivariate Behavioral Research, 18 (January), 85-96.

Green, Paul E. (1975), "Marketing Applications of MDS: Assessment and Outlook," Journal of Marketing, 39 (January), 24-31.

_____ and Frank J. Carmone (1970), Multidimensional Scaling and Related Techniques in Marketing Analysis, Boston, MA: Allyn and Bacon.

Hauser, John R. and Frank S. Koppelman (1979), "Alternative Perceptual Mapping Techniques: Relative Accuracy and Usefulness," Journal of Marketing Research, 16 (November), 495-506.

Jain, Arun K. and Christian Pinson (1976), "The Effects of Order of Presentation of Similarity Judgments on Multidimensional Scaling Results: An Empirical Examination," Journal of Marketing Research, 13 (November), 435-439.

Johnson, Michael D. and Claes Fornell (1987), "The Nature and Methodological Implications of the Cognitive Representation of Products," Journal of Consumer Research, 14 (September), 214-228.

Johnson, Stephen C. (1967), "Hierarchical Clustering Schemes," Psychometrika, 32 (3), 241-254.

Klahr, David (1969), "A Monte Carlo Investigation of the Statistical Significance of Kruskal's Non-metric Scaling Procedure," Psychometrika, 34 (3), 319-330.

Kruskal, J. B. (1964), "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis," Psychometrika, 29 (1), 1-27.

Malholtra, Naresh K. (1987), "Validity and Structural Reliability of Multidimensional Scaling Results," Journal of Marketing Research, 24 (May), 164-173.

_____, Arun K. Jain, and Christian Pinson (1988), "The Robustness of MDS Configurations in the Case of Incomplete Data," Journal of Marketing Research, 25 (February), 95-102.

Moore, William L. and Donald R. Lehmann (1982), "Effects of Usage and Name on Perceptions of New Products," Marketing Science, 1 (Fall), 351-370.

Rosch, Eleanor (1975), "Cognitive Representation of Semantic Categories," Journal of Experimental Psychology: General, 104, 192-233.

Sattath, Shmuel and Amos Tversky (1977), "Additive Similarity Trees," Psychometrika, 42 (3), 319-345.

Shepard, Roger N. (1962), "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function. I and II," Psychometrika, 27 (2), 125-140 and 219-246.

_____ and Phipps Arabie (1979), "Additive Clustering: Representation of Similarities as Combinations of Discrete Overlapping Properties," Psychological Review, 86 (2), 87-123.

Srivastava, Rajendra K, Robert P. Leone, and Allan D. Shocker (1981), "Market Structure Analysis: Hierarchical Clustering of Products Based on Substitution-in-Use," Journal of Marketing, 45 (Summer), 38-48.

Sudman, Seymour and Norman Bradburn (1982), Asking Questions, San Francisco: Jossey-Bass, Inc.

Summers, John O. and David B. MacKay (1976), "On the Validity and Reliability of Direct Similarity Judgments," Journal of Marketing Research, 13 (August), 289-295.

Tversky, Amos (1977), "Features of Similarity," Psychological Review, 84 (4), 327-352.

Weksel, William and Edward E. Ware (1967), "The Reliability and Consistency of Complex Personality Judgments," Multivariate Behavioral Research, 2 (October), 537-541.

ABSTRACT

Similarity scaling often requires subjects to produce such a large number of judgments that subject fatigue may become a problem. Yet it is unclear whether or how respondent fatigue affects similarity judgments. The present study supports the notion that subjects adopt simplified representations and produce less complex judgments as they progress through a similarity rating task.

** Michael D. Johnson is Associate Professor of Marketing and Daniel R. Horne is a doctoral student at the University of Michigan's School of Business Administration, Ann Arbor, Michigan, 48109. Donald R. Lehmann is the George E. Warren Professor of Business, Columbia University, New York, NY, 10027.